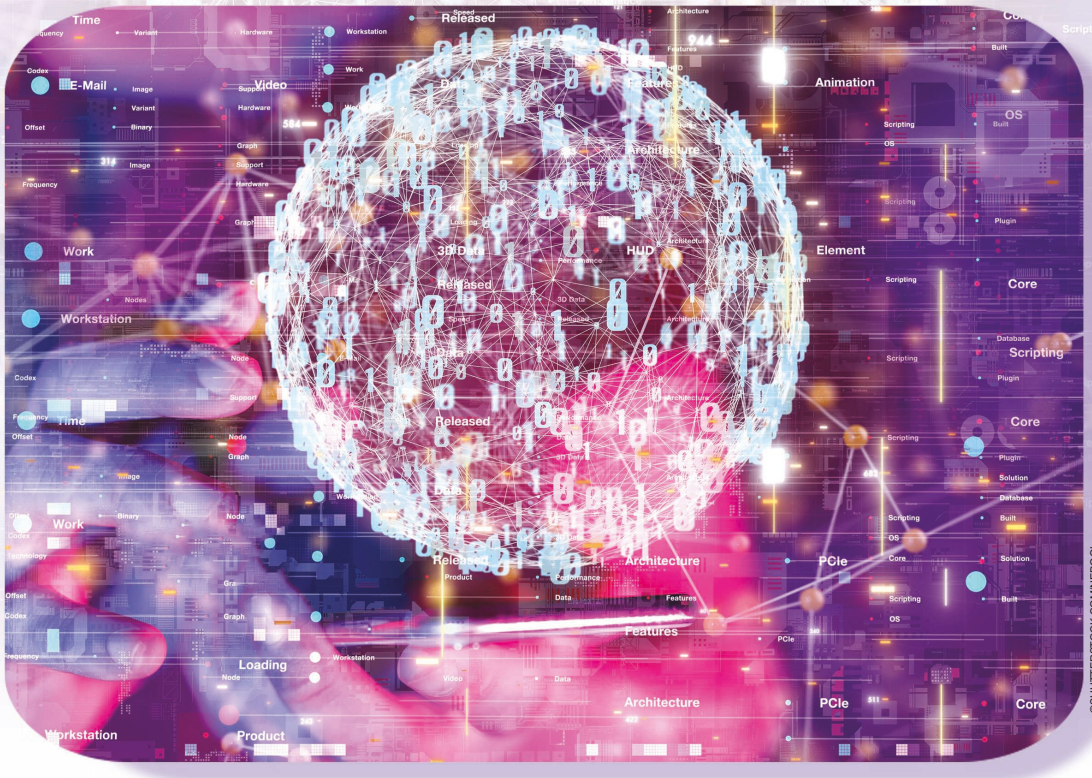


Evolution of Broadband Amplifier Design

Payam Heydari and Mahyar Safiallah

The phrase *distributed amplification* (DA) was first introduced by Ginzton, Hewlett, Jasberg, and Noe in an article published in the August issue of the 1948 Proceedings of the IRE [1]. The authors presented "a new principle in wide-band amplifier design." They then went on to explain that "by an appropriate distribution of ordinary electron tubes along artificial transmission lines, it is possible to obtain amplification over much greater bandwidths than would be possible with ordinary circuits.

Payam Heydari (payam@uci.edu) and Mahyar Safiallah (msafiall@uci.edu) are with the Nanoscale Communication Integrated Circuits Labs, Center for Pervasive Communication and Computing, University of California, Irvine, CA 92617 USA.

Digital Object Identifier 10.1109/MMM.2023.3284729
Date of current version: 6 August 2023

Introduction

The ordinary concept of *maximum bandwidth-gain product* does not apply to this distributed amplifier. The high-frequency limit of the distributed amplifier appears to be determined by the grid-loading effects." Ever since then, distributed amplifiers have been used extensively for broadband wired/wireless applications. Various techniques, examined in different technologies, have been proposed by prior work to improve DA performance parameters such as gain, bandwidth (BW), and power [2], [3], [4], [5], [6], [7], [8], [9]. In works published in 1983–1984, [2], [10] conducted the analysis and design of GaAs field-effect transistor and metal–semiconductor field-effect transistor distributed amplifiers, respectively. Design tradeoffs were established for the type of device, number of gain stages distributed along the input and output lines, gain, impedance level, and line cutoff frequency. Furthermore, the authors detailed a procedure for achieving the maximum gain-bandwidth product. The distributed topology later found its way into CMOS implementation [4] and was used to build several building blocks such as distributed oscillators [4], [11], distributed mixers [12], distributed front ends [13], and distributed active transformers [14]. Meanwhile, research on DA design in other semiconductor technologies continues reporting high-gain and high output power across extremely broad range of frequencies [15], [16], [17].

This article takes a fresh look at distributed amplifiers and provides a step-by-step approach on the basis of the lumped amplifiers and inductive-peaking techniques for bandwidth enhancement. Starting with a resistively loaded conventional common-source amplifier, the work gives a quick overview of the limitations and tight gain-bandwidth tradeoff in the section "Single-Stage Amplifier With its Limitations." The time- and frequency-domain analyses of the amplifier with shunt and shunt-series peaking techniques will be summarized in the sections "Load Resistor Isolation Using Shunt Peaking" and "Load Capacitor and Resistor Isolations Using Shunt-Series Peaking," and several interesting insights will be highlighted. The section "Towards a Distributed Topology" leverages the knowledge laid in the sections "Load Resistor Isolation Using Shunt Peaking" and "Load Capacitor and Resistor Isolations Using Shunt-Series Peaking" and builds the distributed amplifier. This is followed by the section "Improvement of DA's Gain and Linearity," where a few techniques to improve the gain and linearity of the DA will be reviewed. Finally, the last section concludes the article.

Single-Stage Amplifier with Its Limitations

Consider the common-source amplifier in Figure 1(a). As the first step toward understanding the circuit's behavior, the magnitude response, ignoring C_{GD}, is readily derived:

$$\left|\frac{V_{out}}{V_{in}}\right|^2 = \frac{(g_m R_D)^2}{(1+\omega^2 C_D^2 R_D^2)(1+\omega^2 C_G^2 R_G^2)}. \quad (1)$$

As is widely known, with C_{GD} being neglected, the input is isolated from the output node; therefore, each node in the circuit contributes one pole to the circuit's transfer function [Figure 1(b)]. In view of typical values for the source and load impedances, the output pole at $|\omega_{p,out}| = 1/(R_D C_D)$ falls at a lower frequency, and thus, sets the bandwidth. C_D in this circuit is composed of the load and the parasitic capacitors seen looking at the drain terminal. Decreasing R_D or C_D to boost the bandwidth both result in the gain decrease, pointing to a tight tradeoff between gain and bandwidth in this circuit. This limitation does persist when accounting for C_{GD}. Remember that C_{GD} adds a very high-frequency right-half-plane zero (above the transistor's f_T) to the transfer function [18]. Assuming a dominant pole, the 3-dB bandwidth is obtained with the aid of a widely known open-circuit time-constant method [18], i.e.,

$$\omega_{-3dB} = \frac{1}{R_D C_D + R_G C_G + (R_G + R_D + g_m R_G R_D) C_{GD}}. \quad (2)$$

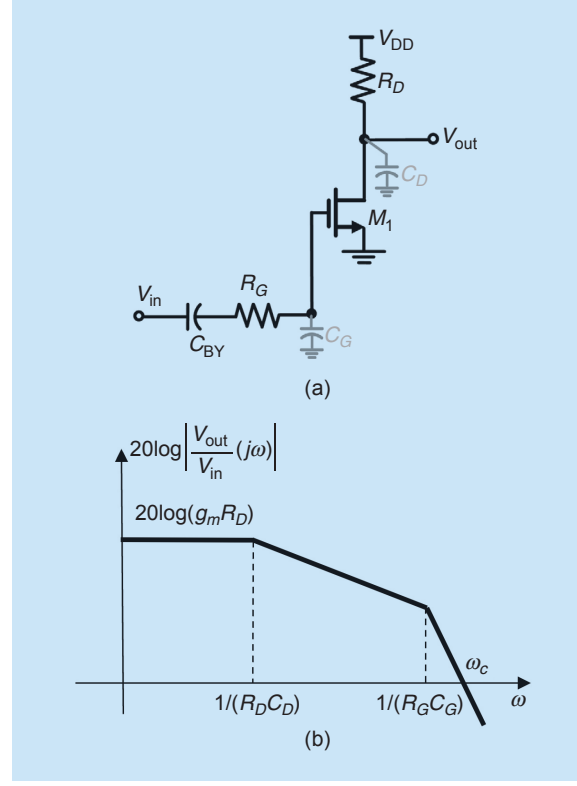


Figure 1. *(a) A conventional common source amplifier. (b) The magnitude response.*

Equation (2) reveals the contribution of each capacitor in the circuit to the bandwidth. In particular, the C_{GD} contribution is seen to get magnified by the Miller effect, which also readily highlights the tight tradeoff between gain and bandwidth. Throughout the analysis, unless otherwise stated, the C_{GD} effect on gain and bandwidth is accounted for using the Miller approximation. The Miller approximation, however, cannot capture the impact of C_{GD} on the group delay and stability of the amplifier. The stability and neutralization techniques to overcome the detrimental effect of C_{GD} on high-frequency amplifiers have been studied in [19].

Load Resistor Isolation Using Shunt Peaking

A commonly known method of relaxing the gain-bandwidth tradeoff and improving the bandwidth is to insert an inductor in series with the resistor, R_D. Called the shunt-peaking technique, the inductor's effect on the circuit of Figure 2(a) can be comprehended by both frequency- and time-domain analyses. Starting with the frequency-domain approach, the circuit magnitude response is readily derived to be as follows:

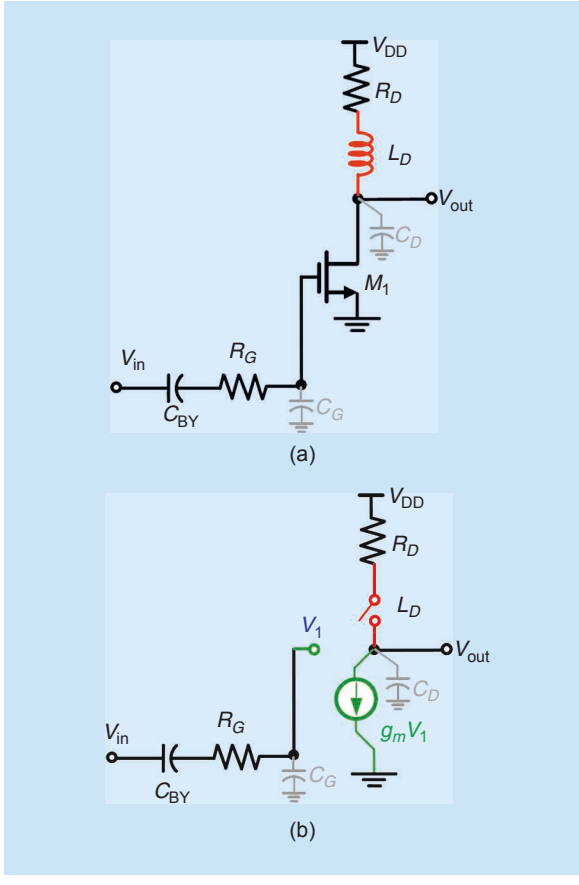


Figure 2. *(a) The circuit incorporating shunt peaking inductor. (b) The equivalent circuit for step-response analysis at time $t = 0^+$.*

$$\left|\frac{V_{out}}{V_{in}}\right|^2 = \frac{(g_m R_D)^2(1+\omega^2 L_D^2/R_D^2)}{(1+\omega^2 C_G^2 R_G^2)[(1-\omega^2 L_D C_D)^2+\omega^2 C_D^2 R_D^2]} \quad (3)$$

Referring to (3), the shunt-peaking inductor will impose two concurrent effects on the magnitude response, both contributing to the bandwidth expansion: 1) It presents a zero to the transfer function. The numerator of the magnitude response thus contains a term that increases with frequency. 2) In the denominator, it introduces a frequency-dependent negative term, $-\omega^2 L_D C_D$, to the second term representing the characteristic polynomial associated with the output node. With a proper choice of L_D, $(1-\omega^2 L_D C_D)$ can stay positive for frequencies up to $1/R_D C_D$. Under this condition, the characteristic polynomial associated with the output node contains a term that decreases with frequency, thus further expanding the bandwidth. The effect of L_D on the amplifier's frequency response is evaluated by simulating a common-source amplifier in a 180-nm CMOS process with a W/L-ratio of 225, operating at a 2-mA of bias current and a 1.8-V supply voltage through a 350-Ω resistor, driving an 80-fF load capacitor. The magnitude response of this amplifier with four inductor values, 0-, 2-, 6-, 10-nH are shown in Figure 3. Larger inductance introduces a larger peaking in the magnitude response.

The time-domain analysis provides a great insight into the circuit operation, as well. From basic circuit theory, the step response of an *RC* circuit shows a transient behavior that can provide information about its bandwidth. In fact, the rise time, t_r, characterizing this behavior is $t_r = 2.2/\omega_{-3dB}$. Suppose that a voltage step is applied at $t = 0$ to the input of the circuit in Figure 2(a). According to the equivalent circuit in Figure 2(b), capturing the circuit behavior at $t = 0$, the inductor cannot carry a step of current at $t = 0^+$, thus acting as an open circuit and cutting off the *R-L* branch from the rest of the circuit. The entire current of the dependent

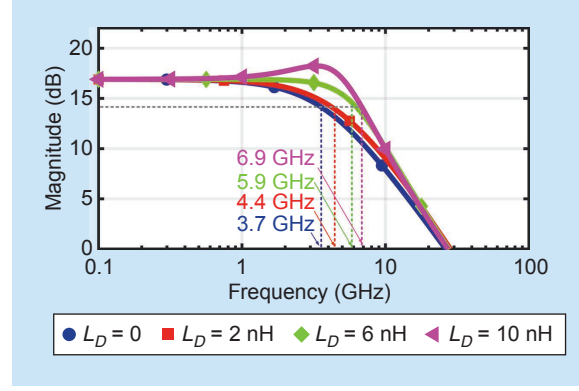


Figure 3. *The magnitude response of a common-source amplifier for four inductor values.*

source, $g_m V_1$, is then available for the capacitor. Being inversely proportional to its current, the capacitor's (dis)charging time is thus shortened compared to the case when the inductor was absent.

Load Capacitor and Resistor Isolations Using Shunt-Series Peaking

How can we further decrease the rise time of the step response? Recall that C_D includes the load as well as the parasitic capacitors. If the load capacitor, C_L, is momentarily detached from the output at $t = 0^+$, the signal transition speed is further accelerated. This can be accomplished by interposing another inductor, L_S, between the load and the transistor's drain, as indicated in Figures 4(a) and (b). At $t = 0^+$, with C_L and R-L branches being open, the entire $g_m V_1$ current will charge only the transistor's own parasitic capacitor, C_P, thereby reducing the rise time.

To evaluate and compare the impacts of shunt- and shunt-series peaking techniques on the time-domain response of a common-source amplifier, we simulated the same common-source amplifier in 180-nm CMOS process with the same aspect-ratio, bias current, and supply voltage as the one in the section "Load Resistor Isolation Using Shunt Peaking." The inductor, L_D, in the shunt-peaked amplifier was 6 nH. The inductors, L_D and L_S, in the shunt-series-peaked amplifier were 3- and 5-nH, respectively. Figure 5(a) shows plots of step responses of three amplifiers subjected to an input step voltage. Referring to the simulation plots, the low-to-high transition time continues to get shortened, as the resistively loaded common-source circuit employs shunt and shunt-series peaking techniques. This shorter transition time trades with another parameter, i.e., delay T_D in rising transition.

As for the spectral analysis of the circuit in Figure 4, rather than investigating the magnitude response of this high-order circuit, a more intuitive approach based on the analysis of resonance behavior would be pursued, here. Depicted in Figure 4(b) is the equivalent circuit to study the circuit's transimpedance, $Z_T = V_{\text{out}}/I_{\text{in}}$. Starting out as resistive load R_D at low frequencies, Z_T will experience the first bump due to the series resonance at $\omega_r = 1/\sqrt{L_S C_L}$. At frequencies above ω_r, the circuit then encounters its first parallel resonance at $\omega_{p1} = (L_S \frac{C_L C_P}{C_L + C_P})^{-0.5}$, where the magnitude rise due to series resonance is partially squared off. Above ω_{p1}, the network becomes capacitive, and the

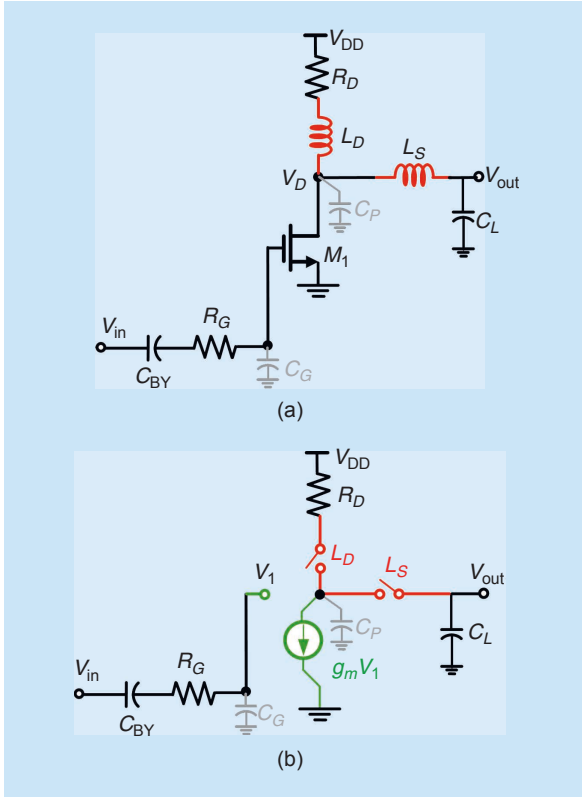


Figure 4. *(a) The circuit incorporating shunt-series peaking inductors. (b) The equivalent circuit for step-response analysis at time $t = 0^+$.*

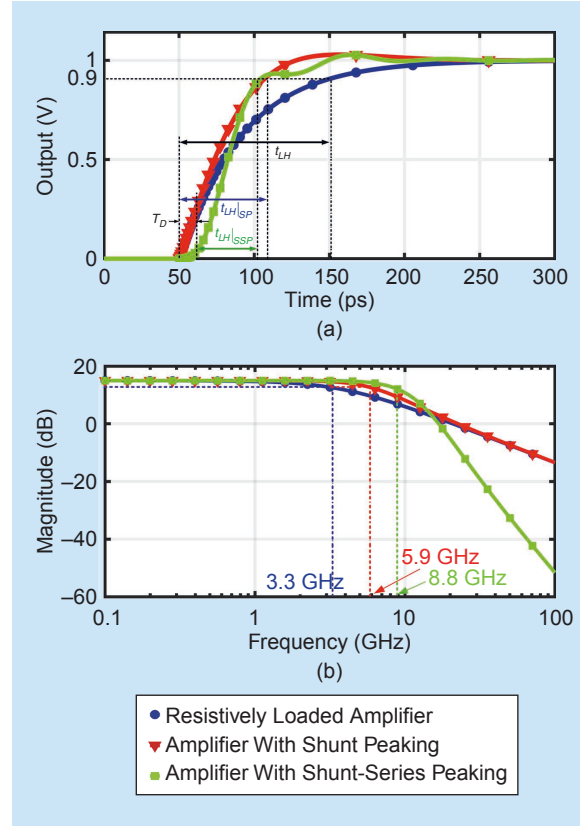


Figure 5. *(a) The step-response of three amplifiers. 1) the amplifier with load resistance, 2) the amplifier employing the shunt peaking, and 3) the amplifier incorporating the shunt-series peaking. (b) Frequency responses of the conventional amplifier and amplifier incorporating shunt- and shunt-series peaking network.*

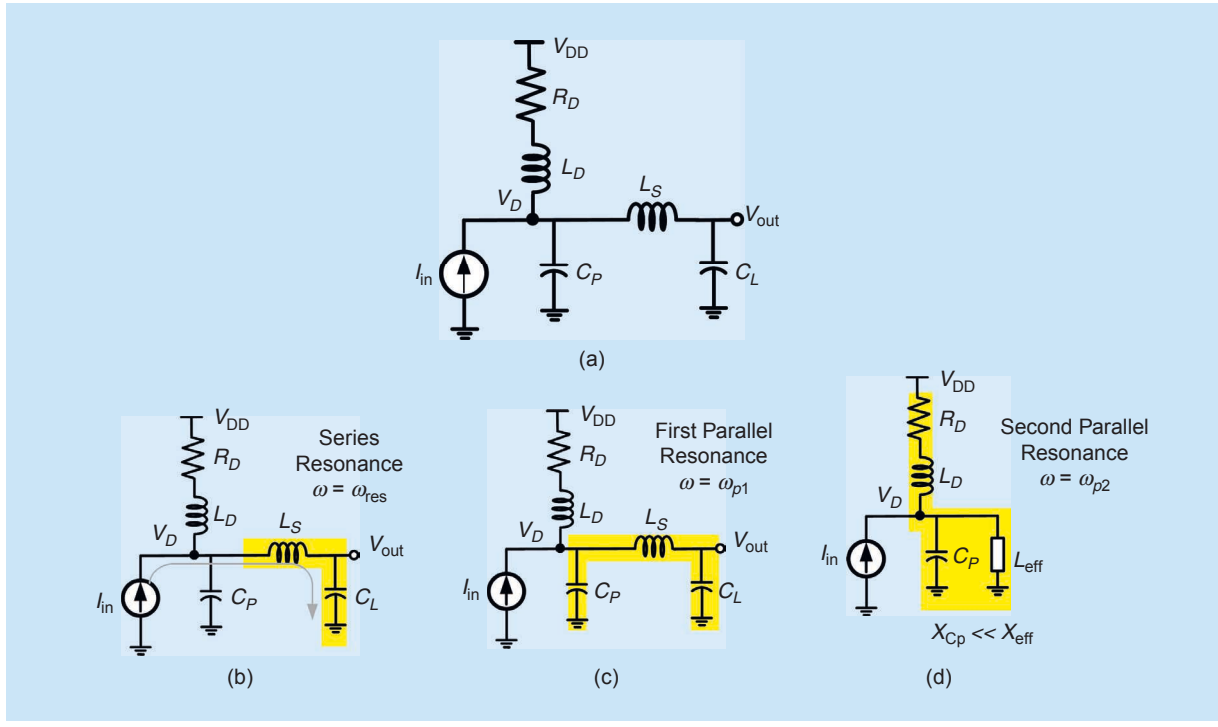


Figure 6. *(a) The triple resonance circuit. (b) The circuit around its series resonance. (c) The circuit behavior around its first parallel resonance. (d) The circuit model at its second parallel resonance.*

frequency response starts falling down with frequency. With further increase in frequency, the second parallel resonance at $\omega_{p2} \approx 1/\sqrt{L_D C_P}$ starts influencing the frequency response, where the inductive behavior below this frequency partially compensates for the roll-off due to the first parallel resonance. At higher frequencies, Z_T will drop because of the capacitive behavior of the circuit above its second parallel resonance (Figure 6). To verify this intuitive analysis, a shunt-series-peaked common source amplifier in 180-nm CMOS with W/L-ratio of 225 is simulated under four different values of the drain resistor, R_D, and the following component values: $L_D = 200$ pH, $L_S = 2.5$ nH, $C_L = 80$ fF. Figure 7 shows the impedance of the shunt-series-peaked amplifier seen from the drain terminal for these distinct R_D values. The impedance plot in Figure 7 clearly shows a dip at the series resonance frequency of 11.2 GHz. A careful inspection also shows the noticeable impact of the first parallel resonance frequency of 19.5 GHz for large drain resistors, while peaking at the second resonance frequency of 56.3 GHz is clearly identified for the 50-Ω load resistor.

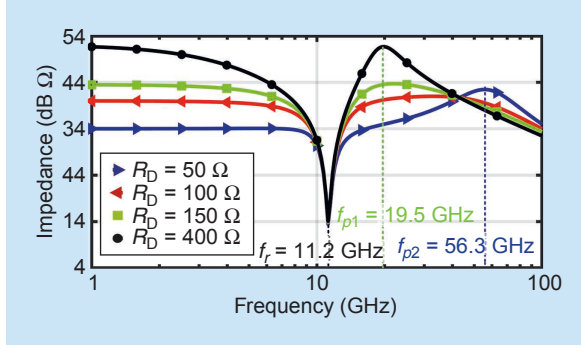


Figure 7. *Driving point impedance of the amplifier incorporating the shunt-series peaking network under three distinct values of the load resistance.*

Toward a Distributed Topology

How can we further decrease the transition time of the step response? Referring to the circuit in Figure 2(a), we learned that inductors L_D and L_S in the series-shunt-peaked amplifier could be used to isolate the resistor and load capacitor so as to decrease the rise time of the step response. Remember that the transition time in a series-shunt-peaked amplifier at $t = 0^+$ is roughly equal to $t_\delta = \frac{C_P V_D}{g_m V_1}$. Given a constant dc power consumption, decreasing the capacitance may provide an effective way of reducing the transition time. To this end, imagine decomposing the device with aspect ratio of W/L into N unit transistors of length L and width $W_k, 1 \le k \le N$ [cf. Figure 8(b)], such that $W = \Sigma_{k=1}^{N} W_k$, and interposing inductors L_u in between the unit transistors [cf. Figure 8(c)]. The current-step arriving at the drain terminal of $M_{1,1}$ starts charging its output parasitic capacitor C_P/N, while parasitic capacitors of other unit cells as well as the load capacitor C_L, are disconnected by the inductor at $t = 0^+$. However, both

the drain current and input/output parasitic capacitors of each cell in Figure 8(c) are scaled by $1/N^{\text{th}}$ of those of M_1. Therefore, absence of supplying additional current per stage, the circuit rise time will not be reduced. Nonetheless, this circuit may provide insight into possible ways of improving the bandwidth.

The use of an inductor in cascade between any two cells introduces "temporal precedence," thus bringing in the notion of "signal propagation." It is inferred that it takes a nonzero time delay, $t_u = \sqrt{L_u C_P/N}$, for the signal to travel from the drain of the first unit cell to that of the second. Let's take a closer look at the signal traveling through the first two stages of the circuit in Figure 8(c). The voltage is applied to the gates of $M_{1,1}$ and $M_{1,2}$, concurrently. The corresponding voltage change at the drain of $M_{1,1}$ in response to this gate voltage should cross past an *LC* network to arrive at the drain of the second cell, thereby arriving later than the voltage that passes through the second stage to appear at this node. This mismatch in arrival times leads to the destructive addition of two copies of the input signal traversing two separate paths.

To equalize this mismatch, we add interstage inductors in between any two adjacent gate terminals in a way that the signals traveling through different paths arrive all at the same time to a specific output node, resulting in the constructive additions of all these signals. In addition, we can move the dc signal feed from the supply to the near-end side of the cascaded *LC* line loading the drain terminals of the transistors to prevent the critical output terminal from loading too many components and their parasitics. This will lead to what is commonly known as a distributed amplifier in Figure 9.

Looking at the circuit in Figure 9, two networks of cascaded *LC* sections are identified, one at the input and the other at the output, where the input and output parasitic capacitors of each MOS device are absorbed into these two *LC* networks. This structure constitutes a cascade of subsections comprising two constant-*k* filters [20], one connecting the gates, and the other, the drains of MOS common-source devices. As was stated in [20], [21], the *image impedance method* efficiently captures the loading effect of subsequent sections on each section and can thus be utilized to analyze the frequency domain behavior of this DA. Referring to the high-frequency model of the DA in Figure 10, both the so-called gate and drain networks are composed of an L-section (in yellow) at the near-end followed by a cascade of $N-1$ π-sections (in peach). The analysis is based on finding the voltage across each C_{GS}, then, calculating the voltage at the drain node of each MOS device in response to this gate voltage, and adding all these voltages (while accounting for the loading effect of subsequent stages) to obtain the overall voltage at

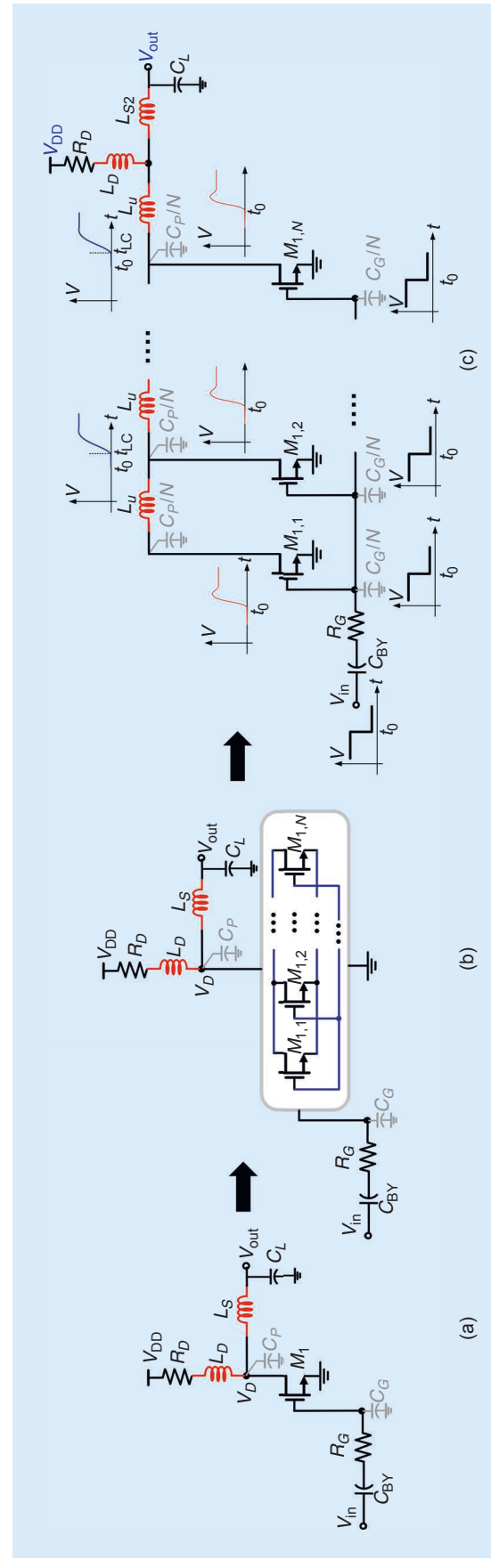


Figure 8. *(a) The triple resonance circuit. (b) Decomposing the transistor into N unit cells.*

the far-end side of the drain node. The voltages V_{GS1} and $V_{GS,k}$ across the capacitors of the L-section and the k*th* π-section in Figure 10 are:

$$V_{GS1} = V_{in}\sqrt{\frac{Z_\pi}{Z_T}}e^{-\theta_G/2} \qquad V_{GS,k} = V_{in}\sqrt{\frac{Z_\pi}{Z_T}}e^{-\left(k-\frac{1}{2}\right)\theta_G} \quad (4)$$

where θ_G in (4), obtained from the image parameter method, defines the propagation factor of the π-section, which is twice that of the L-section. Z_T and Z_π represent the image impedances seen at the input and output ports of the L-section, respectively, which can be found as [21]

$$Z_T = \sqrt{\frac{L_G}{C_G}}\left(1-\frac{\omega^2}{\omega_c^2}\right)^{\frac{1}{2}} \qquad Z_\pi = \sqrt{\frac{L_G}{C_G}}\left(1-\frac{\omega^2}{\omega_c^2}\right)^{-\frac{1}{2}} \quad (5)$$

where $\omega_c = 2/\sqrt{L_G C_G}$ is the gate-line's cutoff frequency. It is noteworthy that Z_π also represents the image impedance of the π-section in Figure 10. The dependent current sources are applied to all tap nodes of the drain line. The voltage at the far-end termination is thus calculated using the superposition principle, i.e.,

$$V_N = -\frac{1}{2}\sqrt{\frac{L_D}{C_D}}e^{-\theta_D/2}\sum_{k=1}^{N} g_m V_{GS,k} e^{-(N-k)\theta_D}. \quad (6)$$

Assuming $\theta_D = \theta_G = \theta$, the voltage and power gains thus become:

$$A_v = -\frac{N g_m}{2\sqrt{1-\omega^2/\omega_c^2}}\sqrt{\frac{L_D}{C_D}}e^{-N\theta} \quad (7)$$

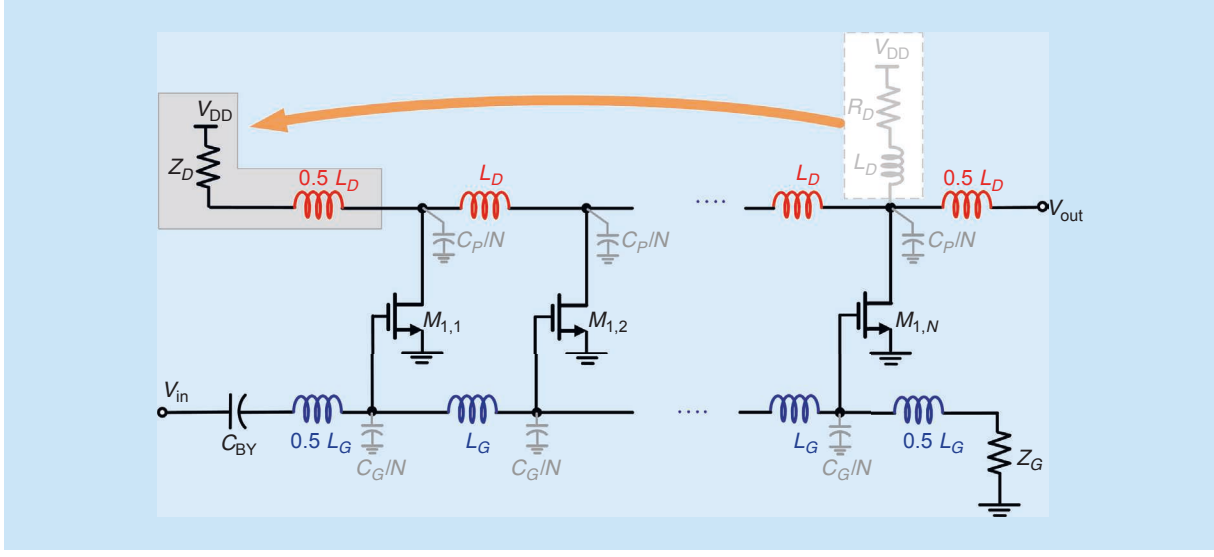


Figure 9. *A DA with common-source stages distributed along the artificial gate and drain LC networks.*

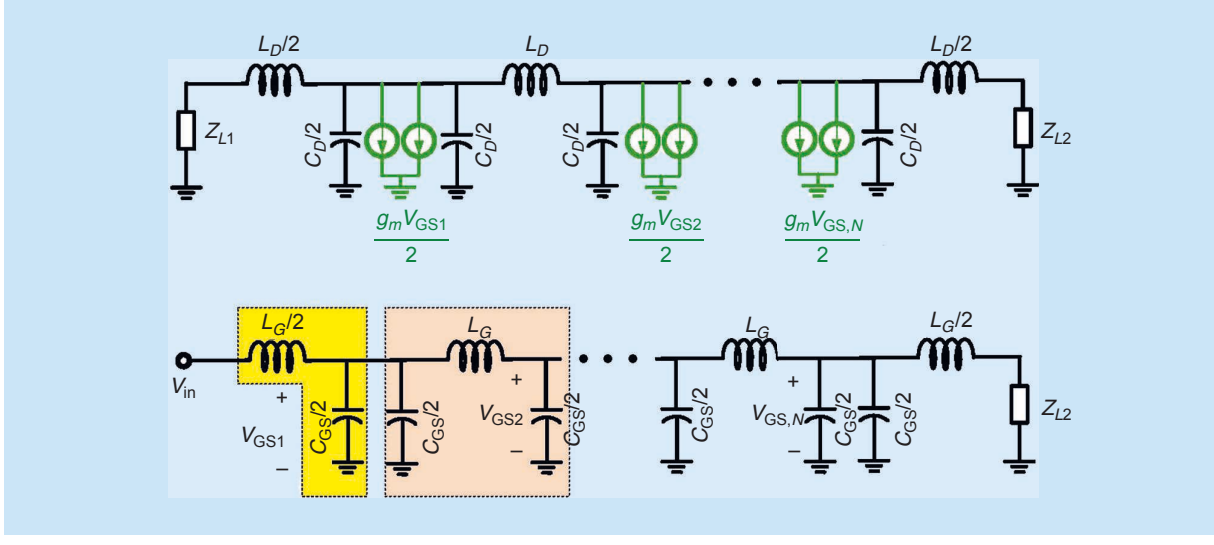


Figure 10. *High-frequency equivalent models of the gate and drain lines.*

$$G_a = \frac{N^2 g_m^2}{4(1-\omega^2/\omega_c^2)}\sqrt{\frac{L_D L_G}{C_D C_G}}\,e^{-2N\theta}. \tag{8}$$

Figure 11(a) shows the input gate voltages and step responses of a 10-stage DA at four distinct tap nodes, where each stage employs a common-source cell with a $g_m = 5$ mA/V distributed along the gate and drain artificial LC networks terminated at 50- and 100-Ω. The gate and drain inductors are 200- and 400-pH, respectively. As expected, the gate voltage of each g_m stage is delayed by $T_G = \sqrt{L_G C_G}$ relative to that of the preceding stage. Referring to Figure 11(b), traveling from the near- to far-end side of the artificial LC line connecting the drain terminals, the rise times of the step responses are progressively shortened. This notion implies that the spectral bandwidth gets progressively bigger as the signal travels from near- to the far-end termination. This observation is, in fact, validated in Figure 12, which demonstrates the frequency responses of the signals appearing at four distinct tap nodes along the drain line. More importantly, one can infer from the behavior of these frequency responses that distributed LC structure of the input/output passive cascaded networks contributes N series-resonance frequencies, which help extend the bandwidth (Figure 12). While the associated dips are distinctly noticeable at the near end, they seem to be flattened due to larger gains at the far-end tap nodes. Finally, the time- and frequency-domain plots in Figures 11(a) and (b) and 12 reveal two attributes of a DA: 1) it trades delay for bandwidth and 2) the transistor's parasitic capacitors are absorbed into the constants of the input/output t-lines. Hence, the circuit bandwidth is expanded and is upper limited by the cutoff frequency of the distributed LC network.

It is noteworthy that the ladder LC networks connecting the gates and drains of the g_m cells would actually model transmission lines (t-lines). The DA can thus be realized using a pair of the gate and drain t-lines in Figure 13, whose voltage gain is expressed by [20]:

$$A_v = -\frac{N g_m}{2}\sqrt{\frac{L_D}{C_D}}\,e^{-N\beta l} \tag{9}$$

where β denotes the propagation constant and $l_D = l_G = l$ is the segment's length. (9) is the same as (7) assuming extremely large ω_c.

The previous analysis was conducted based on the assumption that inductors and t-lines within the DA are lossless and the transistors' output resistances are infinitely large. Accounting for these impairments will lead to degradation of gain and bandwidth.

Improvement of DA's Gain and Linearity

The DA gain is unsatisfactorily small in the presence of passive losses and limited output resistances of common-source transistors, calling for topologies to improve the gain and linearity. The gain improvement in a DA is mainly centered around advancing topologies for the gain stages.

Cascode DA

Multistage g_m stages are used to improve gain, while interstage inductive peaking is employed to compensate for the BW degradation due to interstage poles of these types of stages. Perhaps, the most straightforward improvement is for each stage to employ a cascode topology, indicated in Figure 14, to substantially increase its output resistance [7].

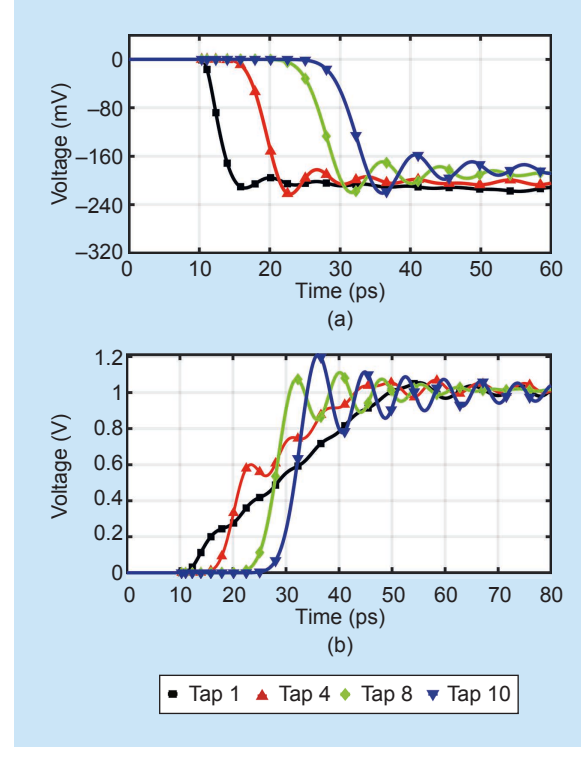


Figure 11. *(a) Gate voltages of g_m cells along the 10-stage DA. (b) Step response of g_m cells along the 10-stage DA.*

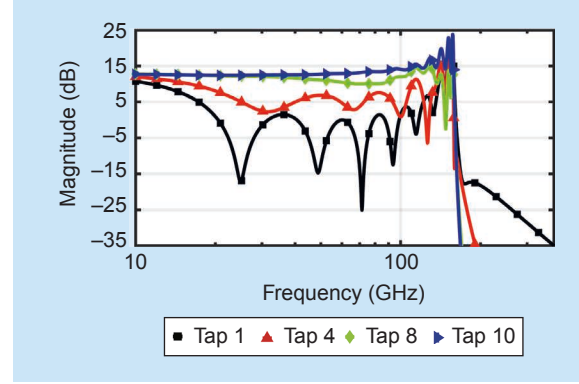


Figure 12. *Frequency response of the 10-stage DA.*

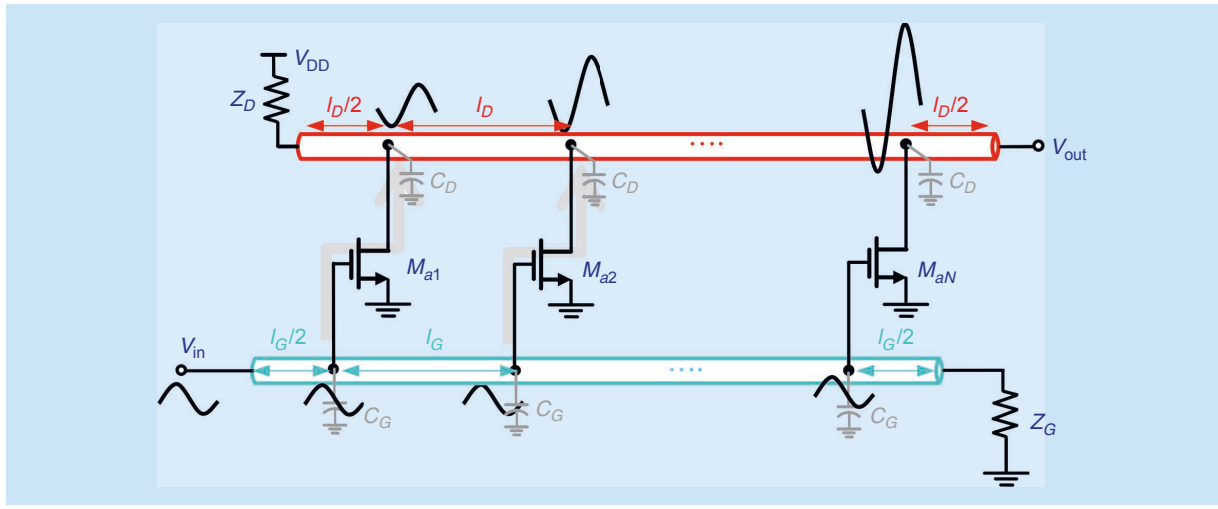


Figure 13. *A DA employing gate and drain t-lines.*

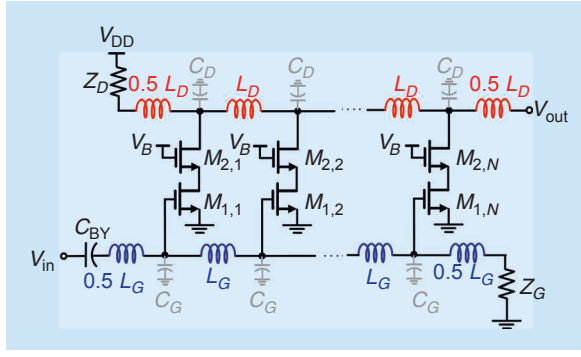


Figure 14. *The Cascode DA.*

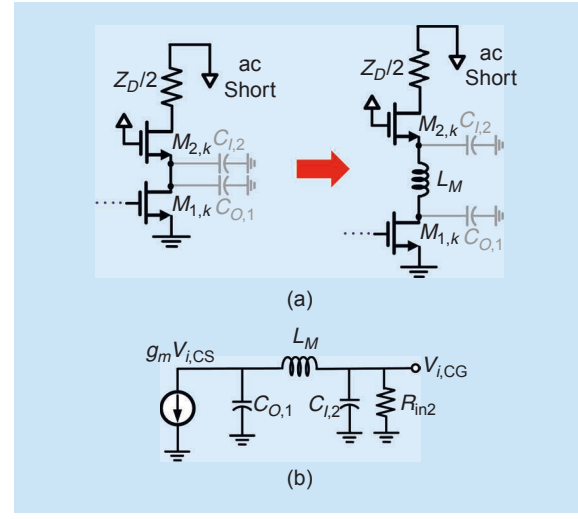


Figure 15. *(a) Evolving from a cascode cell to the one with an interstage inductor between common-source and common-gate devices. (b) The high-frequency model.*

In a cascode DA, the input and output capacitors, C_G and C_D, are absorbed into the structures of input and output t-lines, respectively, and therefore, do not introduce dominant poles. The bottleneck to achieving wide bandwidth is, in this case, the capacitors sitting at the "cascode node," as shown in Figure 15. Intuitively, an interstage series inductor can separate capacitors $C_{O,1}$ and $C_{I,2}$ of the common-source and common-gate transistors, thus increasing the bandwidth. To prove this observation quantitatively, the voltage gain seen from the output of the common source to the input of the common-gate stage within a conventional cascode topology is first derived as follows:

$$\frac{V_{i,CG}}{V_{i,CS}} = -\frac{g_{m1} \times r_{o1} \| R_{in2}}{1 + j\omega (r_{o1} \| R_{in2})(C_{I,2} + C_{O,1})} \tag{10}$$

where $R_{in2} = (r_{o2} + (Z_D/2))/(1 + g_{m2} r_{o2})$ denotes the input resistance of the common-gate stage. The effect of L_M on the transimpedance $Z_T = V_{i,CG}/(g_m V_{i,CS})$ is studied in a similar way as the impact of the series-peaking inductor on a lumped amplifier, illustrated in the section "Load Capacitor and Resistor Isolations Using Shunt-Series Peaking." Specifically, L_M reshapes the frequency response by introducing a series resonance frequency, $\omega_r = (L_M C_{I,2})^{-0.5}$ and a parallel resonance frequency, $\omega_p = (L_M (C_{I,2} C_{O,1}/C_{I,2} + C_{O,1}))^{-0.5}$. Determined by the parallel resonance frequency where the transimpedance experiences a peaking, the bandwidth of the single-stage cascode circuit is expanded to a value approximately equal to the following:

$$BW \approx \omega_p \left(1 + \frac{1}{2Q_{RC}}\right) = \omega_p + \frac{1}{2R_{in2} C_{I,2}}. \tag{11}$$

From another perspective, the common-source devices first start charging their output parasitic and the parasitic capacitors of L_M. Similar to series-peaking

technique, this essentially leads to shorter large-signal transition time and larger bandwidth compared to the case where the inductor is absent.

To verify the aforementioned analytical study, a cascode stage with an identical common-source and common-gate device sizes of 64 μm/0.18 μm and a bias current of 1.1 mA was designed in a 180-nm CMOS. The simulated frequency response of the cascode stage from its input to the common-gate input with and without $L_M = 300$ pH is shown in Figure 16. $C_{I,2}$ and $C_{O,1}$, obtained from the circuit-level simulation, are 114- and 78-fF, respectively. The parallel resonance frequency and the bandwidth obtained from (11) are 42.7 GHz and 51.6 GHz, closely following the corresponding simulated values, 42.7 and 49 GHz, in Figure 16. Based on this finding, a DA comprising N bandwidth-enhanced cascode cells should achieve larger bandwidth than the DA in Figure 14. Three five-stage DAs incorporating gate and drain artificial t-lines with inductors $L_G = 250$ pH and $L_D = 400$ pH were simulated to verify this statement: 1) a DA incorporating common-source cells each with an aspect ratio of 48 μm/0.18 μm, 2) a DA employing cascode cells with common-source and common-gate aspect ratios of 48 μm/0.18 μm and 36 μm/0.18 μm, respectively, and 3) a DA employing the same cascode cells but with series inductor $L_M = 210$ pH. Simulation plots in Figure 17 compare the frequency responses of these three DAs, clearly showing enhancement in both gain and bandwidth when a cascode cell with an interstage BW-enhancing inductor is used.

DAs with Multistage Cells

Figures 18 (a)–(c) demonstrate several multistage g_m cell candidates employed in a DA. Chien and Lu [6] present a two-stage g_m cell in Figure 18(a), which boosts the transconductance from g_{m1} (for a single-stage cell) to $g_{m1}g_{m2}R$. With $g_{m2}R > 1$, gain improvement is guaranteed compared to a DA with single-stage g_m cell used in Figures 9 and 13. In this structure, the BW-enhancement network, L_1-L_2, effectively compensates for the interstage pole associated with the drain of M_1. Nevertheless, the signal power across R is totally wasted, and thus does not participate in signal construction through the output t-line. To circumvent this problem, R is replaced by transistor M_3, as in Figure 18(b). M_3 acts as a common-gate (CG) stage for the new signal path IN - OUT,+ while introducing $1/g_m$ resistive load to the first stage of the two-stage path, IN - OUT,–. Assuming that the output signals OUT,+ and OUT,– are added constructively, the overall G_m of the two-stage cell thus becomes:

$$G_m = g_{m1}\frac{g_{m2}}{g_{m3}} + g_{m1}. \tag{12}$$

M_3's parasitic source capacitance degrades BW compared to Figure 18(a), by lowering the interstage pole of the g_m cell. To mitigate this problem, a network is to be added such that together with the input impedance of the CG transistor constitutes an all-pass filter [9]. This network is synthesized to be a parallel RL network with $R_P = 1/g_{m3}$ and $L_P = R_P^2 C_{I,3}$, as shown in Figure 18(c), where $C_{I,3}$ is the input capacitor of M_3. The equivalent circuit model including the RL network and the impedance seen into the source of M_3 exhibit an all-pass input impedance of $Z_{in} = R_P$ [9].

If two identical g_m cells are connected to differential input lines and their output nodes are cross-coupled as in Figure 18(c), each output of the resulting fully differential g_m cell exhibits the same G_m. The resulting fully differential cell of Figure 18(c) enjoys symmetric design and implementation, and minimal phase and amplitude distortion at its two output terminals. Three five-stage DAs based on the fully differential cell of Figure 18(c) driven by the same 1.8-V supply and a constant $g_{m2}/g_{m3} = 1$ were designed and simulated in 180-nm CMOS process. Figure 19 compares the

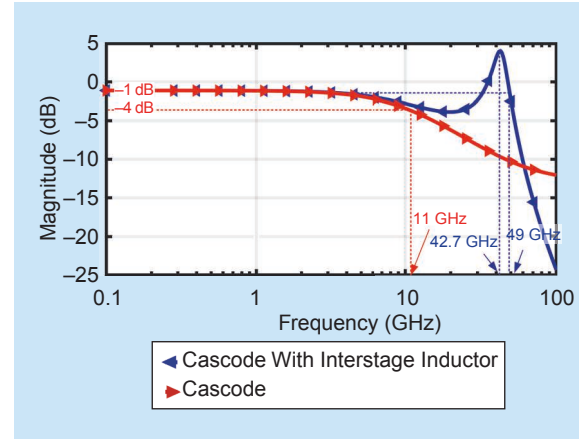


Figure 16. *Frequency response of a single-stage cascode with and without interstage inductor L_M.*

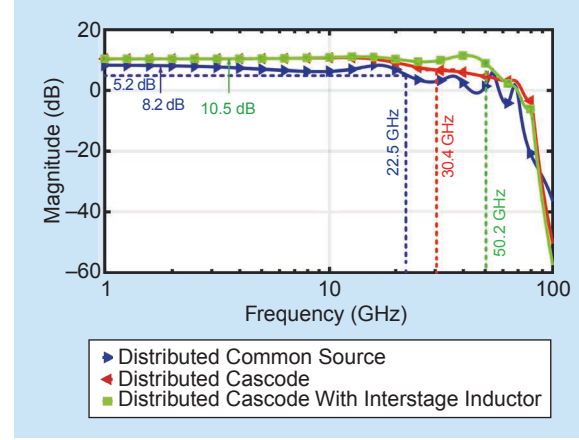


Figure 17. *The DA frequency response.*

frequency responses of these three DAs, and the table in this figure summarizes the device and component values. All these three DAs are terminated with the same 50- and 110-Ω gate and drain resistors, respectively. Moreover, the transistor M_1 employs the same aspect ratio of $48\,\mu\text{m}/0.18\,\mu\text{m}$ and the bias voltage of the common-gate transistors, V_{BIAS}, stays the same at 1.6 V. For approximately the same in-band gain, higher power consumption results in wider bandwidth. All these three DAs exhibit substantially higher GBW compared to a BW-enhanced cascode DA in Figure 17.

Jahanian and Heydari [9] extensively studied issues including stability, power dissipation, noise, and linearity of the DA incorporating the multistage cell in Figure 18(c). Perhaps, the most important issue is the stability concern. Evidently, in the fully differential g_m cell of Figure 18(c), a signal loop is formed due to the cross-coupled connection of the two cells of Figure 18(c). As was proved in [9], this loop entails significant signal attenuation, and therefore, results in an unconditionally stable condition. The existence of this attenuation is evident by a visual inspection of the circuit; going around the loop, the signals arriving at the drain terminals of M_2 and M_5 experience a drop in magnitude by as much as voltage gain to appear at the gates. The DA in Figure 18(c) especially stands out in terms of two important performance parameters, namely, noise and linearity. The gain improvement in the DA of Figure 18(c) results in noise improvement, as well. Second, from the governing dc bias relationship $V_{\text{BIAS}} = V_{\text{GS2}} + V_{\text{GS3}}$ and assuming channel widths of M_1 and M_3 to be identical, one can readily prove that G_m is

$$G_m = \mu_n C_{ox} \frac{\sqrt{W_1 W_3}}{L}(V_{\text{BIAS}} - 2V_{\text{TH}}) \tag{13}$$

where μ_n is the carrier mobility, C_{ox} denotes the oxide capacitance per unit area, and V_{TH} is the threshold

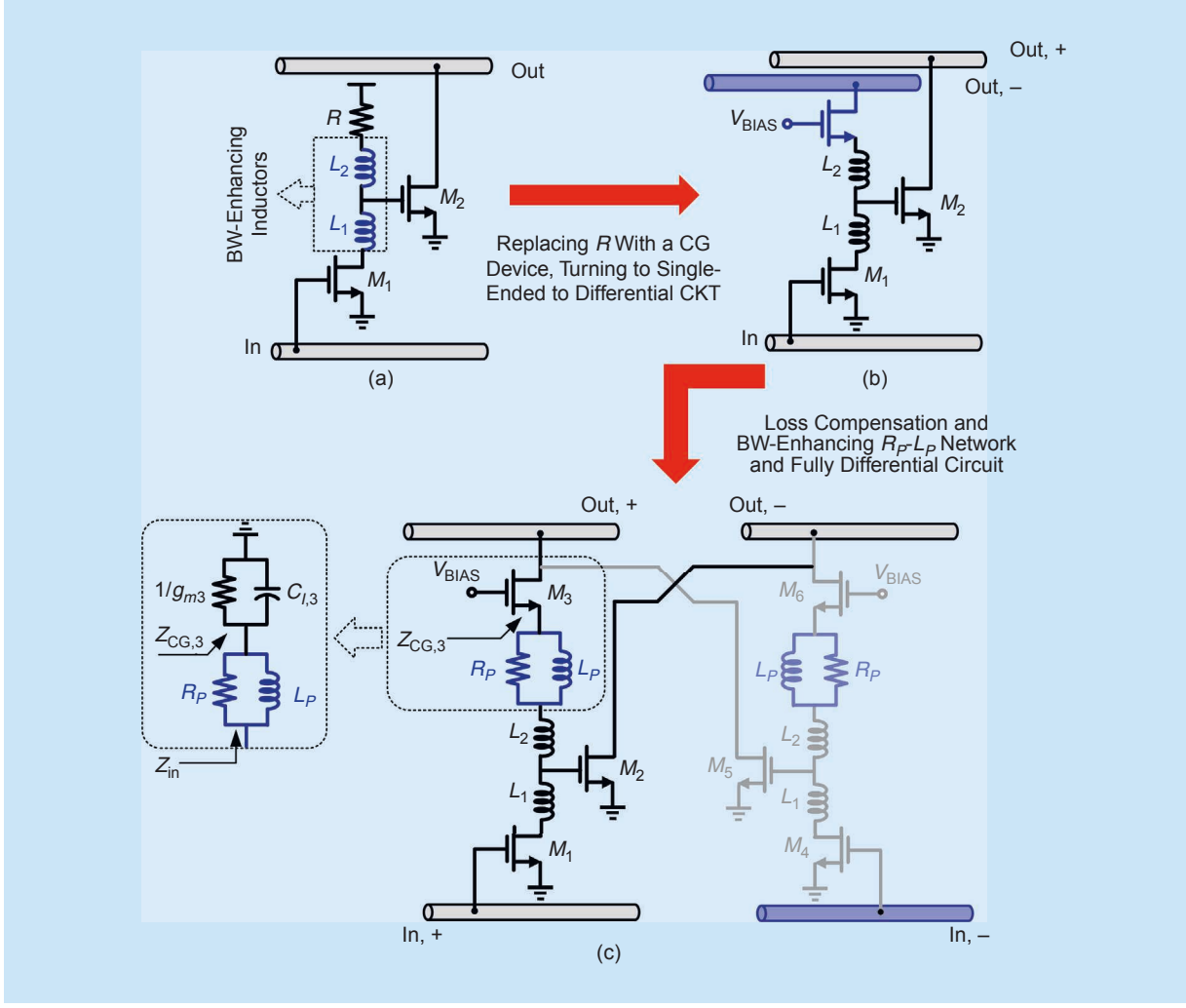
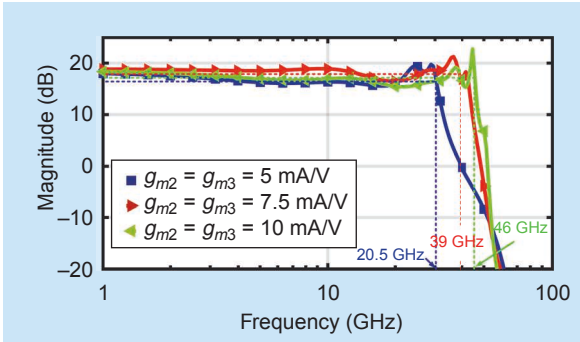


Figure 18. *Multistage DA cell. Evolution from (a) a two-stage CS-CG, to (b) a single-ended to the differential stage, to (c) a fully differential two-stage cell with high gain and high linearity.*



	Diff DA 1	Diff DA 2	Diff DA 3
Number of Stages	5	5	5
$g_{m2,3}$ (mA/V)	5	7.5	10
g_{m1} (mA/V)	16.5	17.19	17.54
$W_1/W_2/W_3$ (μm)	48/32/10	48/21/16	48/24/24
R_D/R_G (Ω)	110/50	110/50	110/50
R_P (Ω)/L_P (pH)	180/50	120/50	87/50
L_D/L_G (pH)	700/320	820/320	600/320
L_1/L_2 (pH)	800/100	400/100	280/50
dc Gain (dB)	17.5	18.6	18
BW (GHz)	20.5	39	46

Figure 19. *The frequency response of three versions of differential DA, and the table summarizing the component values.*

voltage. From (13), it is clear that overall transconductance is constant, and large-signal voltage variation does not affect small signal gain, thereby leading to considerable improvement in linearity.

Conclusion

This article presented a ground-up approach starting from a lumped amplifier and proceeding all the way to a distributed amplifier. The article started with a conventional common-source amplifier and disclosed the limitations and tight tradeoffs. Two widely used bandwidth enhancement techniques, shunt and shunt-series peaking, were reviewed, and implications from both time- and frequency-domain analyses of the amplifier were laid out. Based on this insight, the article took a step-by-step approach to construct the distributed amplifier. Furthermore, this work identified some of the inherent drawbacks and reviewed some techniques that helped improve the gain and linearity of the DA.

Acknowledgment

The authors would like to thank former students in the NCIC Labs including Drs. Amin Jahanian, Aminghasem Safarian, and Lei Zhou for their outstanding contributions to several topics of this work.

References

[1] E. Ginzton, W. Hewlett, J. Jasberg, and J. Noe, "Distributed amplification," *Proc. IRE*, vol. 36, no. 8, pp. 956–969, Aug. 1948, doi: 10.1109/JRPROC.1948.231624.

[2] J. Beyer et al., "MESFET distributed amplifier design guidelines," *IEEE Trans. Microw. Theory Techn.*, vol. 32, no. 3, pp. 268–275, Mar. 1984, doi: 10.1109/TMTT.1984.1132664.

[3] C. S. Aitchison, "The intrinsic noise figure of the MESFET distributed amplifier," *IEEE Trans. Microw. Theory Techn.*, vol. 33, no. 6, pp. 460–466, Jun. 1985, doi: 10.1109/TMTT.1985.1133100.

[4] B. Kleveland et al., "Monolithic CMOS distributed amplifier and oscillator," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC) Tech. Dig. Papers*, 1999, pp. 70–71, doi: 10.1109/ISSCC.1999.759106.

[5] F. Zhang and P. Kinget, "Low-power programmable gain CMOS distributed LNA," *IEEE J. Solid-State Circuits*, vol. 41, no. 6, pp. 1333–1343, Jun. 2006, doi: 10.1109/JSSC.2006.874283.

[6] J.-C. Chien and L.-H. Lu, "40-Gb/s high-gain distributed amplifiers with cascaded gain stages in 180-µm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2715–2725, Dec. 2007, doi: 10.1109/JSSC.2007.908688.

[7] P. Heydari, "Design and analysis of a performance-optimized CMOS UWB distributed LNA," *IEEE J. Solid-State Circuits*, vol. 42, no. 9, pp. 1892–1905, Sep. 2007, doi: 10.1109/JSSC.2007.903046.

[8] A. Arbabian and A. M. Niknejad, "A broadband distributed amplifier with internal feedback providing 660GHz GBW in 90nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC) Tech. Dig. Papers*, 2008, pp. 196–606, doi: 10.1109/ISSCC.2008.4523124.

[9] A. Jahanian and P. Heydari, "A CMOS distributed amplifier with distributed active input balun using GBW and linearity enhancing techniques," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 5, pp. 1331–1341, May 2012, doi: 10.1109/TMTT.2012.2184134.

[10] J. Beyer et al., "Wideband monolithic microwave amplifier study," Dept. Elect. Comput. Eng., Univ. Wisconsin-Madison, Madison, MI, USA, Tech. Rep. ADA136424, 1983.

[11] G. L. G. De Mercey, "A 18GHz rotary traveling wave VCO in CMOS with I/Q outputs," in *Proc. IEEE Eur. Solid-State Circuits Conf.*, 2003, pp. 489–492, doi: 10.1109/ESSCIRC.2003.1257179.

[12] A. Q. Safarian, A. Yazdi, and P. Heydari, "Design and analysis of an ultrawide-band distributed CMOS mixer," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 13, no. 5, pp. 618–629, May 2005, doi: 10.1109/TVLSI.2005.844288.

[13] A. Safarian, L. Zhou, and P. Heydari, "A distributed RF front-end for UWB receivers," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, 2006, pp. 805–808, doi: 10.1109/CICC.2006.321013.

[14] I. Aoki, S. D. Kee, D. B. Rutledge, and A. Hajimiri, "Distributed active transformer-a new power-combining and impedance-transformation technique," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 1, pp. 316–331, Jan. 2002, doi: 10.1109/22.981284.

[15] N. L. Nguyen et al., "A 7–115-GHz distributed amplifier with 24-dBm output power using quadruple-stacked HBT in InP," *IEEE Microw. Wireless Technol. Lett.*, vol. 33, no. 6, pp. 715–718, Jun. 2023, doi: 10.1109/LMWT.2023.3237683.

[16] R. Nikandish, "GaN integrated circuit power amplifiers: Developments and prospects," *IEEE J. Microwaves*, vol. 3, no. 1, pp. 441–452, Jan. 2023, doi: 10.1109/JMW.2022.3221268.

[17] P. T. Nguyen, N. L. Nguyen, A. N. Stameroff, and A.-V. Pham, "A DC-170 GHz InP distributed amplifier using transmission line loss compensation technique," in *Proc. IEEE/MTT-S Int. Microw. Symp.*, Jun. 2022, pp. 841–844, doi: 10.1109/IMS37962.2022.9865541.

[18] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. Hoboken, NJ, USA: Wiley, 2009.

[19] P. Heydari, "Neutralization techniques for high-frequency amplifiers: An overview," *IEEE Solid-State Circuits Mag.*, vol. 9, no. 4, pp. 82–89, Fall 2017, doi: 10.1109/MSSC.2017.2745858.

[20] D. M. Pozar, *Microwave Engineering*. Hoboken, NJ, USA: Wiley, 2011.

[21] T. T. Y. Wong, *Fundamentals of Distributed Amplification*. Norwood, MA, USA: Artech House, 1993.